Keypoints-based background model and foreground pedestrian extraction for future smart cameras

Omar Hamdoun, Fabien Moutarde

Keypoints-based background model and foreground pedestrians extraction for future smart cameras

Omar Hamdoun and Fabien Moutarde
Robotics Laboratory (CAOR), Mines ParisTech, Paris, France
{Omar.Hamdoun, Fabien.Moutarde}@mines-paristech.fr

Abstract— **In this paper, we present a method for background modeling using only keypoints, and detection of foreground moving pedestrians using background keypoints substraction followed by adaBoost classification of foreground keypoints. A first experimental evaluation shows very promising detection performances in real-time.**

Keywords: video-surveillance, background substraction, interest points.

I. INTRODUCTION

Traditionally, multi-camera video-surveillance systems are monitored by human operators, and usually saved to tapes for later forensic use. The increase in the number of cameras in those surveillance systems overloaded both the number of operators and the storage devices, and made it impossible to ensure proper monitoring of sensitive areas for long time periods. In order to filter out redundant information and reduce the response time to important events, development of Intelligent Surveillance Assistance Systems (ISAS) for helping the human operators is essential. In the case of large multi-camera system, it may be difficult to transmit and process all video streams in one single central computer or even in several local ones, so our vision is that future smart cameras will extract useful primitives such as interest points, and ISAS systems should thus be designed for processing streams of such higher-level information rather than videos.

The first step in nearly every visual surveillance system is detection of moving objects. Object detection aims at segmenting regions corresponding to moving objects from the rest of the image. Subsequent processes such as classification and tracking are greatly dependent on it. The process of object detection usually involves background modelling and motion segmentation. Due to dynamic changes in natural scenes such as sudden illumination, weather changes and repetitive motions that cause clutter, object detection is a difficult problem to solve reliably. Frequently used techniques for moving object detection are background substraction, statistical methods, temporal differencing and optical flow.

Our aim is to adapt this processing chain to the case where, instead of full video streams, only interest point descriptors from various smart cameras are available. We present here a method combining background modeling and substraction using interest points, and adaBoost classification for selection of pedestrian-specific interest points. First experiments show this approach can robustly detect moving foreground pedestrians with low false alarm rate.

II. METHOD

In this section we detail the algorithmic choices made in our detection approach. Our method can be divided in two main stages: keypoints background modeling, then classification filtering.

A. *Description of the background keypoints model*

We use the "keypoints" functions available in the Camellia (http://camellia.sourceforge.net) image processing library. These Camellia keypoints detection and characterization functions implement a very fast variant of SURF [2]. The latter itself is a very efficient method inspired by the more classical and widely spread interest point detector and descriptor SIFT [3]. Another main difference between Camellia keypoints and SURF is that the Camellia implementation uses integer-only computations – even for the scale interpolation –, which makes it even faster than SURF, and particularly well-suited for potential embedding in smart camera hardware, contrary to SIFT and SURF whose existing implementations both make extensive use of floating point computations.

The background is modelled by keypoints, each with its descriptor, and an associated distance threshold (equal to the maximum descriptor distance for similar neighbouring keypoints between consecutive frames). In our implementation we selected an array V containing the interest points of N consecutive images: $V_i(\phi)$ is the descriptor of an interest point in approximate location ϕ in the i^{th} image. The initial background model for a keypoint at approximate location ϕ is $[m(\phi), D(\phi)]$ obtained with following equations:

$$\begin{cases} m_0(\phi) = V_0(\phi) \\ \mathrm{D}_0(\phi) = 0 \end{cases} \quad (1a) \qquad \begin{cases} m_{i+1}(\phi) = \alpha V_i(\phi) + (1-\alpha) m_i(\phi) \\ \mathrm{D}_{i+1}(\phi) = max\left(\mathrm{D}_i(\phi), \left|V_{i+1}(\phi) - V_i(\phi)\right|\right) \end{cases} \quad (1b)$$

where α is the learning rate. In our experiments, we use $\alpha = 0.15$. Keypoints matching is done with a very fast Best Bin First (BBF) descriptor search in a KD-tree [4] containing all interest points of all frames. When no match is found for a keypoint, we add it to the KD-tree.

Classification: each keypoint is classified as background or foreground as follows: given the values of $m(\phi)$ and $D(\phi)$, a keypoint $Kp(\phi)$ is considered as foreground if :

$$\left|Kp(\phi) - m(\phi)\right| > D(\phi) \quad (2)$$

After this foreground filtering, the new keypoints are also used to continuously update the background model. The new keypoints in the image are added to the model, while keypoints that have not been present for a while are removed. This allows for objects that quickly move in and out of the scene to be considered foreground, while new objects that come into the scene and remain stationary will automatically be slowly blended into the background.

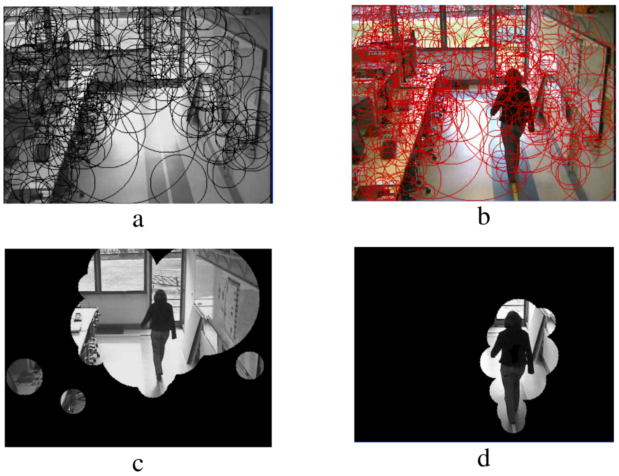

Figure 1. (a) Background keypoints model; (b) All keypoints in current image; (c) Result of foreground selection; (d) Foreground pedestrian area obtained after further adaBoost filtering

The main problem of interest points is their instability at smaller scales, because they are more susceptible to changes in lighting and camera noise, and are often on edges. In addition, most of them probably describe background clutter or ambiguous objects, making them difficult to match. These unstable keypoints cause some false alarms (i.e. errors in foreground selection, see figure 1c). To overcome this problem, we applied a second filtering step based on keypoints classification.

B. *Filtering using adaBoost classification of keypoints*

Many techniques have been proposed for visual object detection and classification (see e.g. [1] for a review of some of the state-of-the-art methods for pedestrian detection, which is the most challenging, and [6] for real-time detection using adaBoost). Recently, it was also shown [7] that boosting based on "keypoint presence" weak classifiers can be used for object recognition, and that keypoints characteristic of a given object class can be determined.

We therefore decided, for our keypoint classification filter, to use adaBoost for recognizing pedestrian-specific keypoints. We collected keypoints from a set of pedestrian images (50.000 keypoints for 600 pedestrians) and non-pedestrian samples (150.000 keypoints). We then applied a particular implementation of boosting, realBoost, which builds a CART (Classification and Regression Tree) [5] by combining weak classifiers testing independantly one vector coordinate. CART is a tree graph, with leaves representing the classification result and nodes representing some predicate. The branches of the tree are marked true or false. The classification process in case of decision trees is a process of tree traverse.

The obtained classifier can determine for each keypoint if it is or not a "pedestrian keypoint". It is used as a second filter, applied only to the foreground keypoints obtained after substraction of background keypoints described in II.A. This second filtering allows to remove the remaining "false foreground keypoints". The final result is shown in figure 1d, where a mask was created by placing circles at each keypoint location, sized according to its scale. Note that some morphological filtering (opening and closure) is also applied to clean up the mask.

III. CONCLUSIONS AND PERSPECTIVES

Considering the probable emergence of smart cameras computing high-level features such as interest points, the main objectives of our approach is to develop an algorithm for building a background model and performing foreground moving object detection using only interest points (without the video frames). There are currently very few algorithms that address this issue. Our approach to solving this problem is unique in that it uses keypoints (~ SURF interest points) both for constructing the background model and for object detection. This keypoints-based approach adds robustness to the solution, because it can be used to reduce the noise that is often encountered in color based detection.

More thorough evaluations have to be done and are currently under progress, including on other video corpus, in order to assess detection reliability during a large period of time. Also, our detection scheme will soon be integrated in a global video-surveillance processing, including our recently proposed method for inter-camera pedestrian re-identification using keypoints [8].

REFERENCES

[1] Munder S. and Gavrila D. M., "An Experimental Study on Pedestrian Classification". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.28, nr 11, pp. 1863-1868, 2006.

[2] Bay H., Tuytelaars T. & Gool, L. V. "SURF:Speeded Up Robust Features", *Proceedings of the 9th European Conference on Computer Vision (ECCV'2006)*, Springer LNCS volume 3951, part 1, pp 404--417, 2006.

[3] Lowe, D. "Distinctive Image Features from Scale-Invariant Keypoints" *International Journal of Computer Vision, Vol. 60*, pp. 91-110*, Springer,* 2004.

[4] Beis, J. & Lowe, D., "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces", In *Proc. 1997 IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1000-1006, Puerto Rico*,* 1997.

[5] Friedman, J.; Hastie, T. & Tibshirani, R. additive logistic regression: A statistical view of boosting *Annals of statistics, Institute of Mathematical Statistics,* 2000, 337-3

[6] Viola, P. & Jones, M., "Rapid object detection using a boosted cascade of simple features" *Proc. CVPR,* 2001*, 1*, 511-518

[7] Bdiri T., Moutarde F., and Steux B., "Visual object categorization with new keypoint-based adaBoost features, *Proc. of IEEE Intelligent Vehicles symposium, XiAn (China), June 2009.*

[8] Hamdoun, O.; Moutarde, F.; Stanciulescu, B. & Steux, B. "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences, *2008.. Second ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2008), Stanford (USA), september* 2008.